

**Ridge Regression and Generalized Maximum Entropy:
an improved version of the Ridge-GME parameter estimator¹**

Pedro Macedo

CIDMA – Center for Research and Development in Mathematics and Applications,
Department of Mathematics, University of Aveiro,
Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
pmacedo@ua.pt

In this paper, the Ridge-GME parameter estimator, which combines Ridge Regression and Generalized Maximum Entropy, is improved in order to eliminate the subjectivity in the analysis of the ridge trace. A serious concern with the visual inspection of the ridge trace to define the supports for the parameters in the Ridge-GME parameter estimator is the misinterpretation of some ridge traces, in particular where some of them are very close to the axes. A simulation study and two empirical applications are used to illustrate the performance of the improved estimator. A MATLAB code is provided as supplementary material.

Keywords: generalized maximum entropy, ridge regression, shrinkage estimation

MSC 2010: 62J07, 94A17

¹Final version of this paper is published in **Communications in Statistics – Simulation and Computation**, 46(5), 3527-3539, 2017 (<http://dx.doi.org/10.1080/03610918.2015.1096378>).

1 Introduction

It is well-known that under fairly regularity conditions, linear regression becomes a powerful statistical tool. In practice, however, some of these conditions are usually not satisfied and regression models become ill-posed, implying that the application of traditional estimation methods, such as the Ordinary Least Squares (OLS) or the Maximum Likelihood (ML) estimators, may lead to non-unique or highly unstable solutions. One of the major concerns in multiple linear regression is the collinearity² problem, which is responsible for inflating the variance associated with the regression coefficients estimates³ and, in general, to affect the signs of the estimates, as well as statistical inference. Common examples arise from many fields including astronomy, chemistry, medicine and economics, where it is necessary to make the best possible predictions with limited and noisy information.

Various methods, including Ridge Regression, Principal Component Regression, Partial Least Squares Regression, Continuum Regression, Lasso, Elastic Net, Least Angle Regression and Generalized Maximum Entropy are well suited to cope with collinearity problems; see, among others, Brown (1994), Ciavolino and Al-Nasser (2009), Efron et al. (2004), Golan (2006), Hoerl and Kennard (1970b), Satici and Demirhan (2012), Stone and Brooks (1990), Tibshirani (1996) and Zou and Hastie (2005) for reviews. However, it is interesting to note that, despite some more recent approaches, Ridge Regression continues to play a key role in regression models affected by collinearity and outperforms other competitors in many cases; e.g., Hastie et al. (2009), Maronna (2011), McDonald (2009), Zou and Hastie (2005). The importance of Ridge Regression is discussed by McDonald (2009) that analyzed the number of publications in the *Technometrics*, the *Journal of the American Statistical Association*, the *Communications in Statistics – Theory and Methods*, and the *Communications in Statistics – Simulation and Computation*. Approximately 320 articles related to Ridge Regression has been published in these four scientific journals since the seventies.⁴

²Naturally, this notion is not used here in a literal sense; e.g., Belsley et al. (2004, pp. 85–98).

³Small sample size (micronumerosity) is also responsible for the inflation of variance.

⁴This number of publications results from a recent update based on the study of McDonald (2009).

The Ridge Regression introduced by Hoerl and Kennard (1970b) is an estimation procedure to handle collinearity without removing variables from the regression model. By adding a small non-negative constant (tuning, ridge or shrinkage parameter) to the diagonal of the correlation matrix of the explanatory variables, it is possible to reduce the variance of the OLS estimator through the introduction of some bias. Although the resulting estimators are biased, the biases are small enough for these estimators to be substantially more precise than the unbiased estimators. Regarding this usual concern between the use of biased and unbiased estimators, it is important to note that OLS and ML may also be biased; see, for example, Ryan (2009, p. 466) and Greene (2008, p. 487). Furthermore, and considering other criticisms on Ridge Regression, it is worth to mention that ML estimator may not be unique and it is attractive mainly due to the large-sample properties. Moreover, considering β as the vector of unknown parameters in a regression model, OLS considers that $\beta'\beta$ is unbounded. Naturally, $\beta'\beta$ is unknown but should be finite, which means that assuming an upper bound on $\beta'\beta$ and applying OLS subject to this restriction, the general form of the Ridge Regression is obtained (Ryan, 2009, p. 470). Thus, it is difficult to argue against Ridge Regression and the usual criticisms in choosing biased estimators may not be always totally fair.

The challenge in Ridge Regression remains on the selection of the ridge parameter. One straightforward approach is based on simply plotting the coefficients against several possible values for the ridge parameter and inspecting the resulting traces; e.g., Hoerl and Kennard (1970b,a), Zhang and McDonald (2005). In addition to this graphical procedure, several formal methods have been proposed to estimate the ridge parameter. The literature on methods for choosing the ridge parameter is massive and includes, among many others, Alkhamisi and Shukur (2007), Gibbons (1981), Golub et al. (1979), Hoerl and Kennard (1970b,a), Hoerl et al. (1975), Khalaf and Shukur (2005), Kibria (2003), McDonald and Galarneau (1975) and Muniz and Kibria (2009).

Recently, based on the Generalized Maximum Entropy (GME) estimator, the Ridge-GME parameter estimator (Macedo et al., 2010) appears in the literature as one interesting

approach. The use of the GME estimator tries to solve the problem of the subjective selection of the tuning parameter based on the visual inspection of the ridge trace, and the Ridge Regression approach provides information for the selection of the supports for the parameters in the regression model, which is the main drawback of the GME estimator. Thus, by circumventing criticisms and difficulties of traditional Ridge Regression and GME estimation, the Ridge-GME parameter estimator seems to be a perfect combination between the estimators that highlights the best of both.

However, the Ridge-GME parameter estimator requires the use of subjective information from the analysis of the ridge trace, which represents the main drawback of the estimator. In this paper, the Ridge-GME parameter estimator is developed so that no subjective information from the visual inspection of the ridge trace is needed. Furthermore, a MATLAB code with the improved Ridge-GME parameter estimator is provided.

The remainder of the paper is laid out as follows: in Section 2, a brief overview on Ridge Regression and GME estimation is presented. In Section 3, the improvement on the Ridge-GME parameter estimator is discussed. A simulation study is provided in Section 4. In Section 5, two empirical applications are accomplished. Finally, some concluding remarks are given in Section 6.

2 Ridge Regression and Generalized Maximum Entropy

Consider the multiple linear regression model defined by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

where \mathbf{y} denotes a $(N \times 1)$ vector of noisy observations, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of unknown parameters, \mathbf{X} is a known $(N \times K)$ matrix of explanatory variables and \mathbf{u} is a $(N \times 1)$ vector of random disturbances (errors), usually assumed to have a conditional expected value of zero and representing spherical disturbances, i.e, $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ and $E[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \sigma^2\mathbf{I}$, where

\mathbf{I} is a $(N \times N)$ identity matrix and σ^2 is the error variance.

The OLS and the Ridge Regression estimators of $\boldsymbol{\beta}$ are well-known.

Definition 2.1. *The Ordinary Least Squares (OLS) estimator of $\boldsymbol{\beta}$ in model (1) is given by*

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2)$$

Definition 2.2. *The Ridge Regression estimator of $\boldsymbol{\beta}$ in model (1) takes the form*

$$\hat{\boldsymbol{\beta}}_{ridge} = (\mathbf{X}'\mathbf{X} + \eta\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}, \quad (3)$$

where $\eta \geq 0$ denotes the ridge parameter and \mathbf{I} is a $(K \times K)$ identity matrix.

Definition 2.2 represents the biased estimator proposed by Hoerl and Kennard (1970b). Note that when $\eta \rightarrow 0$, the Ridge Regression estimator approaches the OLS estimator whereas the Ridge Regression estimator approaches the zero vector when $\eta \rightarrow \infty$. Thus, a trade-off between variance and bias is needed. Hoerl and Kennard (1970b, pp. 62–63) proved that the Ridge Regression estimator is superior to the OLS estimator (in a Mean Squared Error sense) for a range of values of η , say $0 < \eta < \sigma^2/\alpha_{\max}^2$, where α_{\max}^2 is the largest squared value from a vector $\boldsymbol{\alpha}$ that depends on $\boldsymbol{\beta}$ (considering model (1) in the canonical form). Thus, is the problem of choosing η solved? Unfortunately the answer is no, because σ^2 and $\boldsymbol{\beta}$ are unknown.

Definition 2.3. *The Mean Squared Error (MSE) of the Ridge Regression estimator of $\boldsymbol{\beta}$ in model (1) is given by*

$$MSE(\hat{\boldsymbol{\beta}}_{ridge}) = \sigma^2 \sum_{k=1}^K \frac{\lambda_k}{(\lambda_k + \eta)^2} + \eta^2 \boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + \eta\mathbf{I})^{-2}\boldsymbol{\beta}, \quad (4)$$

where λ_k are the eigenvalues of the $\mathbf{X}'\mathbf{X}$ matrix in correlation form.

It is interesting to note that the variance of $\hat{\boldsymbol{\beta}}_{ridge}$ decreases as η increases, whereas the bias increases with η . Due to its biasedness, the Ridge Regression estimator is superior to the OLS estimator (in a MSE sense) if the reduction in the variance is greater than the increase

of the squared bias. Since the range of values for which the Ridge Regression estimator is superior to the OLS estimator depends on the unknown parameters $\boldsymbol{\beta}$ and σ^2 , the challenge is to select an estimate of η such that the Ridge Regression estimator has a smaller MSE than the OLS estimator.

To better understand the drawbacks of the OLS estimator under collinearity, consider the expected squared distance between $\hat{\boldsymbol{\beta}}_{OLS}$ and $\boldsymbol{\beta}$, defined as

$$E \left[(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta})' (\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \right]. \quad (5)$$

Considering $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K > 0$ as the ordered eigenvalues of the $\mathbf{X}'\mathbf{X}$ matrix in correlation form, it follows that

$$E \left[(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta})' (\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \right] = \sigma^2 \sum_{k=1}^K \frac{1}{\lambda_k} \quad (6)$$

and

$$E \left[\hat{\boldsymbol{\beta}}_{OLS}' \hat{\boldsymbol{\beta}}_{OLS} \right] = \boldsymbol{\beta}' \boldsymbol{\beta} + \sigma^2 \sum_{k=1}^K \frac{1}{\lambda_k}. \quad (7)$$

Thus, as λ_K becomes smaller (leading to an increase in collinearity), the vector $\hat{\boldsymbol{\beta}}_{OLS}$ can be expected to be farther from the vector $\boldsymbol{\beta}$; e.g., Hoerl and Kennard (1970a).

As noted by Golan et al. (1996), statistical data are frequently limited and affected by collinearity implying that the associated statistical models may be ill-posed, unless simplifying assumptions/procedures are imposed to generate seemingly well-posed statistical models that can be estimated with traditional statistical tools. Giving heed to this problem, Golan et al. (1996) generalized the Maximum Entropy formalism (Jaynes, 2003) to linear inverse problems with noise, expressed in model (1). The idea is to treat each β_k as a discrete random variable with a compact support and $2 \leq M < \infty$ possible outcomes, and each u_n as a finite and discrete random variable with $2 \leq J < \infty$ possible outcomes. Assuming that both the unknown parameters and the unknown error terms may be bounded a priori, the linear model (1) can be presented as

$$\mathbf{y} = \mathbf{XZp} + \mathbf{Vw}, \quad (8)$$

where

$$\boldsymbol{\beta} = \mathbf{Z}\mathbf{p} = \begin{bmatrix} \mathbf{z}'_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}'_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}'_K \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_K \end{bmatrix}, \quad (9)$$

with \mathbf{Z} a $(K \times KM)$ matrix of support values and \mathbf{p} a $(KM \times 1)$ vector of unknown probabilities, and

$$\mathbf{u} = \mathbf{V}\mathbf{w} = \begin{bmatrix} \mathbf{v}'_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{v}'_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{v}'_N \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix}, \quad (10)$$

with \mathbf{V} a $(N \times NJ)$ matrix of support values and \mathbf{w} a $(NJ \times 1)$ vector of unknown probabilities.

Definition 2.4. *For the linear regression model specified in (1), the Generalized Maximum Entropy (GME) estimator is given by*

$$\operatorname{argmax}_{\mathbf{p}, \mathbf{w}} \{ -\mathbf{p}' \ln \mathbf{p} - \mathbf{w}' \ln \mathbf{w} \}, \quad (11)$$

subject to the model constraint

$$\mathbf{y} = \mathbf{X}\mathbf{Z}\mathbf{p} + \mathbf{V}\mathbf{w}, \quad (12)$$

and the additivity constraints for \mathbf{p} and \mathbf{w} , respectively,

$$\begin{aligned} \mathbf{1}_K &= (\mathbf{I}_K \otimes \mathbf{1}'_M)\mathbf{p}, \\ \mathbf{1}_N &= (\mathbf{I}_N \otimes \mathbf{1}'_J)\mathbf{w}, \end{aligned} \quad (13)$$

where \otimes represents the Kronecker product, $\mathbf{1}$ is a column vector of ones with a specific dimension, \mathbf{I} is an identity matrix with a specific dimension and, as defined in (9) and (10), \mathbf{Z} and \mathbf{V} are the matrices of supports, and $\mathbf{p} > \mathbf{0}$ and $\mathbf{w} > \mathbf{0}$ are probability vectors to be estimated.

The GME estimator generates the optimal probability vectors $\hat{\mathbf{p}}$ and $\hat{\mathbf{w}}$ that can be used to form point estimates of the unknown parameters and the unknown random errors through

the reparameterizations (9) and (10), respectively. As noted by Golan et al. (1996), since the objective function (11) is strictly concave in the interior of the additivity constraint set, a unique solution for the GME estimator is guaranteed if the intersection of the model and the additivity constraint sets is non-empty.

The supports in matrices \mathbf{Z} and \mathbf{V} are defined as being closed and bounded intervals within which each parameter or error is restricted to lie, implying that researchers need to provide exogenous information (which, unfortunately, it is not always available). This is considered the main weakness of the GME estimator; see, for example, Caputo and Paris (2008) for further details. Golan et al. (1996) discuss these issues in the case of minimal prior information: for the unknown parameters, the authors recommend the use of wide bounds (this is naturally subjective) for the supports in \mathbf{Z} , without extreme risk consequences; for the unknown errors, the authors suggest the use of the three-sigma rule with a sample scale parameter. The number of points, M and J , in the supports is less controversial and are usually used in the literature between 3 and 7 points, since there is likely no significant improvement in the estimation with more points in the supports. The three-sigma rule, considering the standard deviation of the noisy observations and $J = 3$ points in the supports, is adopted in the improved Ridge-GME parameter estimator discussed in Section 3.

3 Improved Ridge-GME parameter estimator

The Ridge-GME parameter estimator was introduced by Macedo et al. (2010) and was adapted for a jackknife procedure by Erdugan and Akdeniz (2012). The basic idea underlying the Ridge-GME parameter estimator is to combine the ridge trace and the GME estimator. The two key issues are: how can the information provided by the ridge trace be used without making a subjective selection of an estimate of η ?, and how can the ridge trace and the GME estimator be efficiently combined? In answering to these questions, Macedo et al. (2010) illustrate how the Ridge-GME parameter estimator can be implemented. Moreover, based on the results from the simulation studies conducted by Macedo et al. (2010) and Erdugan

and Akdeniz (2012), it seems reasonable to state that, in the case of regression models with small samples sizes affected by collinearity, the Ridge-GME parameter estimator has a good performance and it may be recommended to practitioners.

However, the original Ridge-GME parameter estimator has two drawbacks: it requires subjective information from visual inspection of the ridge trace, namely the definition of the ridge interval (a set of possible values for the ridge parameter) and it requires the supports for the parameters of the regression model. These drawbacks are overcome with the improved version of the Ridge-GME parameter estimator discussed next.

First, considering the definition of the ridge interval, since it is widely accepted in the literature that an estimate of the ridge parameter must be selected within the interval $[0, 1]$, the improved Ridge-GME parameter estimator considers this interval by default, with 1000 possible values for the ridge parameter. Naturally, the user has always the option to define another ridge interval as a subset of the default interval $[0, 1]$.⁵

Second, considering the definition of the supports for the parameters, Figure 1 illustrates the improved Ridge-GME parameter estimator, where the subjective choice of the supports for the parameters based on the visual inspection of the ridge trace is eliminated. In the original Ridge-GME parameter estimator, the ridge trace provides “guidelines” for the selection of the supports for the model regression parameters, which leads to subjective choices for the supports.⁶ A serious concern with the visual inspection of the ridge trace to define the supports for the parameters is the misinterpretation of some ridge traces, in particular where some of them overlap (or are very close to) the axes. For example, this difficulty occurs in both empirical applications discussed in Section 5: in Figure 2 one trace exceeds the value -2 and in Figure 3 one trace exceeds the value 62. In both cases, the limits of these traces are very difficult to obtain only by visual inspection of the ridge traces.

⁵In this case, the ridGME code (the MATLAB code provided as supplementary material) presents the ridge trace to help the user in the selection of the ridge interval.

⁶For example, based on the same ridge trace for the Portland cement model, different supports for the parameters are considered by Macedo et al. (2010) and Erdugan and Akdeniz (2012).

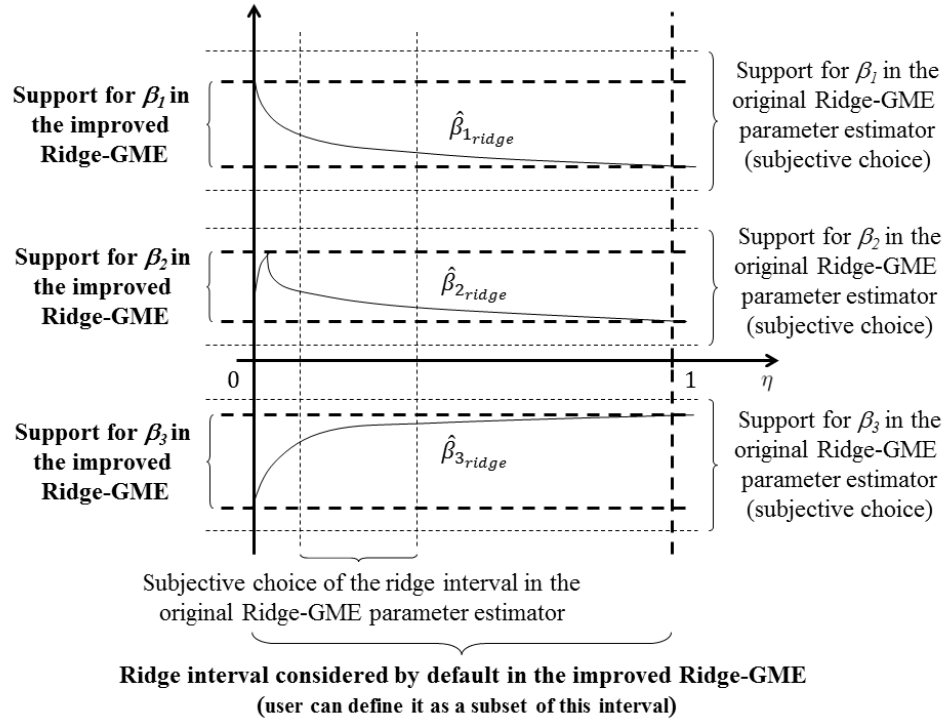


Figure 1: Ridge interval and supports in the improved Ridge-GME parameter estimator.

Definition 3.1. Considering the vector $\hat{\mathbf{p}}$ obtained through the GME estimator, where the matrix \mathbf{Z} is a $(K \times 5K)$ matrix of support values with limits given by

$$\mathbf{z}_k = \left[\min \left\{ \hat{\beta}_{k_{ridge}(\eta)} \right\}, \max \left\{ \hat{\beta}_{k_{ridge}(\eta)} \right\} \right], \quad (14)$$

for all the 1000 possible values of η in the ridge interval, $[\eta_1, \eta_{1000}]$, which is $[0, 1]$ by default, the improved Ridge-GME parameter estimator is given by

$$\hat{\eta} = \underset{\eta}{\operatorname{argmin}} \left\| \mathbf{Z}\hat{\mathbf{p}} - (\mathbf{X}'\mathbf{X} + \eta\mathbf{I})^{-1} \mathbf{X}'\mathbf{y} \right\|_{\infty} \quad (15)$$

subject to the specified ridge interval.

With the improved Ridge-GME parameter estimator, given the ridge interval, $[\eta_1, \eta_{1000}]$, the supports for the parameters are no longer subjectively defined.

4 Simulation study

A similar simulation study to the one conducted by Macedo et al. (2010) is provided in this section to compare the original Ridge-GME parameter estimator with the improved version proposed in Definition 3.1. Results from the OLS estimator are also provided for comparison.

The five explanatory variables are generated through

$$x_{ij} = (1 - \alpha^2)^{\frac{1}{2}} z_{ij} + \alpha z_{i6}, \quad (16)$$

$i = 1, 2, \dots, n$, $j = 1, 2, \dots, 5$, where z_{ij} are independent standard normal pseudo-random numbers and α is specified so that the correlation between any two explanatory variables is given by α^2 ; e.g., Gibbons (1981), Kibria (2003) and McDonald and Galarneau (1975). Four different combinations of sample sizes are considered, namely $n = 10, 20, 50, 100$. By choosing the true coefficient vector β as the normalized eigenvector corresponding to the largest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix, the n observations for the dependent variable are obtained by

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + u_i, \quad (17)$$

$i = 1, 2, \dots, n$, where u_i are independent normal pseudo-random numbers with zero mean and variance σ^2 . Three different values for σ and five different values for α are tested, namely $\sigma = 0.5, 1.0, 1.5$ and $\alpha = 0.750, 0.900, 0.950, 0.975, 0.999$.

The ridge interval for the original Ridge-GME parameter estimator is defined as $[0, 1]$ and the GME estimator is performed using different supports, namely $[-5, 5]$ and $[-10, 10]$, for the five parameters of the model, illustrating the possible subjectivity in the selection of the supports. The 3σ rule is used to define the supports for the error component, where σ represents the empirical standard deviation of the noisy observations. The number of support points used for the unknown parameters and the error component is 5 and 3, respectively. The improved version of the Ridge-GME parameter estimator is implemented according to Definition 3.1, where the ridge interval is considered as $[0, 1]$ by default.

Since the variables were standardized, the estimated standardized coefficients are transformed back to the original model and for the 1000 trials performed, the mean squared

error loss (MSEL), with $\text{SEL}(\hat{\beta}) = \left(\|\hat{\beta} - \beta\|_2\right)^2$, is the measure used to evaluate the performance of the three estimators.

Table 1: MSEL for OLS and Ridge Regression ($n = 10$ and $n = 20$).

α	OLS	Ridge with improved		OLS	Ridge with improved		
		Ridge with Ridge-GME	Ridge-GME		Ridge with Ridge-GME	Ridge-GME	
		$n = 10$				$n = 20$	
		$\sigma = 0.5$				$\sigma = 0.5$	
0.750	0.4560	0.2601	0.0892	0.1569	0.1249	0.0978	
0.900	1.8412	0.3757	0.0717	0.2823	0.1951	0.0818	
0.950	2.0110	0.2408	0.0439	0.5979	0.1428	0.0362	
0.975	4.1578	0.3749	0.0718	1.5466	0.3829	0.0453	
0.999	69.7136	0.1243	0.9997	39.0571	0.0112	0.6029	
		$\sigma = 1.0$				$\sigma = 1.0$	
0.750	3.0610	1.9030	0.3725	0.7170	0.5088	0.1197	
0.900	6.7313	1.0928	0.1525	2.0817	0.7244	0.0874	
0.950	8.6043	1.1746	0.1476	3.3028	0.7251	0.0752	
0.975	20.1567	1.3260	0.3398	7.4598	0.9681	0.1302	
0.999	1028.7500	3.4956	11.3115	102.4354	0.1227	1.5318	
		$\sigma = 1.5$				$\sigma = 1.5$	
0.750	3.9923	2.9679	0.6539	1.8092	1.2537	0.2134	
0.900	8.8877	5.4661	0.7010	5.4705	1.5241	0.1424	
0.950	31.6143	3.5083	0.5184	6.8170	2.3432	0.1633	
0.975	70.5696	3.2625	0.9404	13.2649	1.2399	0.2213	
0.999	763.5422	7.7537	9.4897	276.4247	0.5046	3.5355	

Table 1 and Table 2 presents the results for OLS and Ridge Regression with the original Ridge-GME parameter estimator, considering the support $[-10, 10]$ for the five parameters

Table 2: MSEL for OLS and Ridge Regression ($n = 50$ and $n = 100$).

α	Ridge with improved			Ridge with improved		
	OLS	Ridge with Ridge-GME	Ridge-GME	OLS	Ridge with Ridge-GME	Ridge-GME
$n = 50$				$n = 100$		
$\sigma = 0.5$				$\sigma = 0.5$		
0.750	0.0622	0.0428	0.0572	0.0399	0.0259	0.0396
0.900	0.1217	0.0616	0.0395	0.0549	0.0277	0.0371
0.950	0.2433	0.0811	0.0348	0.1111	0.0369	0.0331
0.975	0.4971	0.0577	0.0299	0.1949	0.0347	0.0307
0.999	14.2421	0.0183	0.2319	5.2173	0.0064	0.0967
$\sigma = 1.0$				$\sigma = 1.0$		
0.750	0.2586	0.1899	0.0868	0.0950	0.0734	0.0710
0.900	0.5576	0.3140	0.0594	0.2272	0.1289	0.0466
0.950	1.0015	0.3124	0.0410	0.4943	0.1576	0.0365
0.975	2.2038	0.3022	0.0416	0.8207	0.1535	0.0310
0.999	48.4125	0.0373	0.8342	21.9658	0.0204	0.3562
$\sigma = 1.5$				$\sigma = 1.5$		
0.750	0.4922	0.3411	0.0861	0.2495	0.1783	0.0699
0.900	1.2959	0.6825	0.0764	0.5530	0.2898	0.0549
0.950	2.3680	0.7513	0.0632	0.9839	0.2881	0.0372
0.975	3.7130	0.8399	0.0677	1.9774	0.3450	0.0406
0.999	78.4112	0.1075	1.1941	46.1627	0.0520	0.7002

of the model, and with the improved version of the estimator proposed in Definition 3.1. As expected, in terms of MSEL, both Ridge Regression procedures outperform the OLS estimator in all the scenarios. Concerning the two Ridge Regression procedures, in the 60 simulations performed, the Ridge Regression estimator with the improved version of

the Ridge-GME parameter estimator outperforms the Ridge Regression estimator with the original version in 45 scenarios, in terms of MSEL. In general, the worst results with the improved version of the Ridge-GME parameter estimator are obtained only for $\alpha = 0.999$, regardless the value of n and σ , which represents the cases of extreme severe collinearity.

This simulation study reveals that the performance of the original Ridge-GME parameter estimator depends on the supports for the parameters. For example, considering some supports of higher amplitude than with $[-10, 10]$, the Ridge Regression estimator with the improved version of the Ridge-GME parameter estimator outperforms the Ridge Regression estimator with the original version in all the 60 simulations. However, for example, considering the support $[-5, 5]$ for the five parameters of the model, the Ridge Regression estimator with the improved version of the Ridge-GME parameter estimator outperforms the Ridge Regression estimator with the original version only in 38 of the 60 simulations.⁷

These results highlight the concerns with the visual inspection of the ridge trace to define the supports for the parameters in the original Ridge-GME parameter estimator. With the improved version proposed in Definition 3.1, the supports for the parameters are no longer arbitrarily defined and seem to be appropriately specified in most of the cases considered in this simulation study.

5 Empirical applications

In this section, the well-known acetylene (Marquardt and Snee, 1975) and the Portland cement (Woods et al., 1932) models are used to illustrate the performance of the improved Ridge-GME parameter estimator. These data sets have received considerable attention in the literature; e.g., Montgomery et al. (2006), Hald (1952), Kaçiranlar et al. (1999), Liu (2003), Muniz and Kibria (2009), Sakallıoğlu and Kaçiranlar (2008).

⁷Results not reported here due to space limitations are provided upon request to the author.

5.1 Acetylene model

The response variable in the acetylene model is the percentage of conversion of n -heptane to acetylene (\mathbf{y}) and the three explanatory variables are the reactor temperature (\mathbf{x}_1), the ratio of H_2 to n -heptane (\mathbf{x}_2) and the contact time (\mathbf{x}_3). The data set is composed by 16 observations. Following Montgomery et al. (2006, p. 329), the full quadratic model usually considered for the acetylene data is

$$\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_3 + \beta_4\mathbf{x}_1\mathbf{x}_2 + \beta_5\mathbf{x}_1\mathbf{x}_3 + \beta_6\mathbf{x}_2\mathbf{x}_3 + \beta_7\mathbf{x}_1^2 + \beta_8\mathbf{x}_2^2 + \beta_9\mathbf{x}_3^2 + \mathbf{u}. \quad (18)$$

The model is affected by severe collinearity ($\text{cond}_2 \mathbf{X} \approx 208$)⁸ which means that OLS estimator should not be used. The ridge trace in Figure 2 illustrates the lack of stability of the OLS estimates and the presence of severe collinearity, i.e., large changes in the coefficients for small values of the ridge parameter.

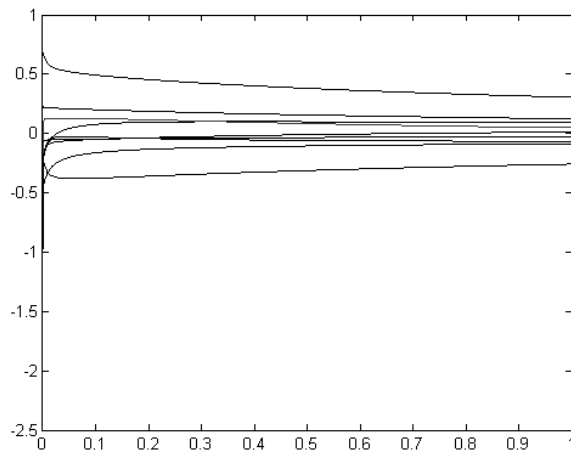


Figure 2: Ridge trace for the acetylene model.

From visual inspection of the ridge trace in Figure 2, Montgomery et al. (2006, p. 347) consider that reasonable coefficient stability is achieved in the region established by $0.008 <$

⁸Ratio of the largest singular value of the design matrix, \mathbf{X} , with the smallest singular value (condition number). The regressors and the response are centered and scaled.

$\eta < 0.064$ and, without a rule to choose the ridge parameter, the authors select $\eta = 0.032$ for the Ridge Regression model, obtaining a residual mean square value of, approximately, 9.4×10^{-4} .

Considering the original Ridge-GME parameter estimator (Macedo et al., 2010), from visual inspection of the ridge trace in Figure 2, the ridge interval can be subjectively defined, for example, by $\eta \in [0, 0.2]$, and different supports can be subjectively specified for the parameters of the model, corresponding to different estimates for the ridge parameter (different combinations of supports were tested in this model, considering equal supports for all the parameters and also considering specific supports for each parameter). The residual mean square values obtained varies between 6.1×10^{-4} and 4.0×10^{-3} , approximately.

Finally, considering the improved Ridge-GME parameter estimator without any subjective information (considering option 1 in the ridGME code, i.e., considering $\eta \in [0, 1]$ by default), the estimated ridge parameter is $\eta = 0.001$ and the corresponding residual mean square value is, approximately, 4.7×10^{-4} . It is interesting to note that considering, for example, the specific ridge interval defined by Montgomery et al. (2006, p. 347), the estimated ridge parameter is $\eta = 0.0181$ (considering option 2 in the ridGME code) and the corresponding residual mean square value is, approximately, 7.7×10^{-4} . In this case, the main advantage of the improved Ridge-GME parameter estimator is that no kind of judgment is required to select an estimate of η in the interval $[0.008, 0.064]$.

Thus, with lower residual mean square values and, most importantly, with no need for any judgement in the interpretation of the ridge trace, the improved Ridge-GME parameter estimator seems to be an attractive approach.

5.2 Portland cement model

The response variable in this model is the heat evolved per gram of cement (\mathbf{y}) and the four explanatory variables are the amounts of tricalcium aluminate (\mathbf{x}_1), tricalcium silicate (\mathbf{x}_2), tetracalcium aluminoferrite (\mathbf{x}_3) and β -dicalcium silicate (\mathbf{x}_4). The linear model without intercept presented by Woods et al. (1932) does not suffer from collinearity because

$\text{cond}_2 \mathbf{X} \approx 21$, where \mathbf{X} is the matrix of the explanatory variables. However, the linear model with intercept defined as

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \beta_4 x_{n4} + u_n, \quad (19)$$

$n = 1, 2, \dots, 13$, is affected by severe collinearity since $\text{cond}_2 \mathbf{X} \approx 6056$, with \mathbf{X} representing the matrix of the explanatory variables with the first column of ones. As noted by Liu (2003), this dramatic change in cond_2 is explained by the fact that the sum of each row in the original \mathbf{X} matrix is approximately equal to 100 (the explanatory variables are presented in rounded percentages) and, thus, the model (19) is affected by severe collinearity.

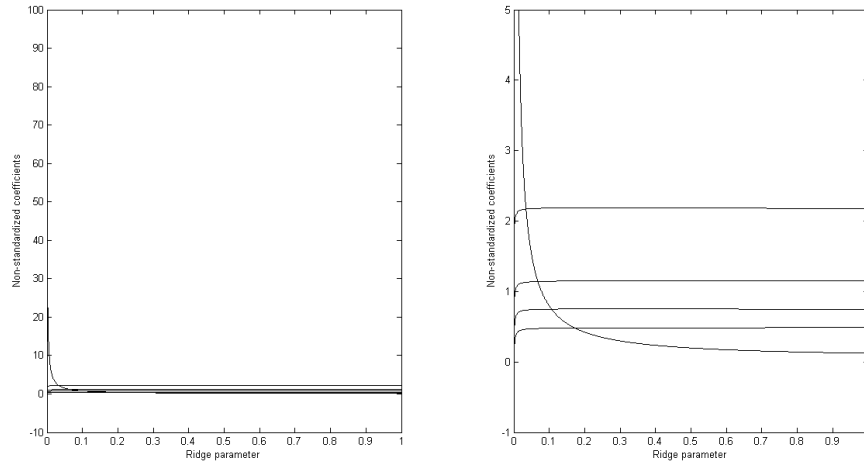


Figure 3: Ridge trace for the Portland cement model.

The ridge trace in Figure 3 illustrates the lack of stability of the OLS estimates and the presence of severe collinearity. From visual inspection of the ridge trace in Figure 3, the ridge interval was defined as $\eta \in [0, 0.3]$ by Macedo et al. (2010). Table 3 shows the MSE for the OLS and the Ridge Regression estimator based on the improved Ridge-GME parameter estimator without any subjective information (option 1 in the ridGME code) and using the same ridge interval defined by Macedo et al. (2010) (option 2 in the ridGME code), the generalized cross-validation (GCV) estimator by Golub et al. (1979), the HK estimator

proposed by Hoerl and Kennard (1970b,a), the HKB estimator by Hoerl et al. (1975), the KS estimator by Khalaf and Shukur (2005), and the KM4, KM5 and KM6 estimators by Muniz and Kibria (2009).

Table 3: MSE for different estimators in the Portland cement model.

	MSE
OLS	4912
Ridge with KM5	3882
Ridge with KM4	3879
Ridge with KM6	3876
Ridge with GCV	3859
Ridge with HKB	2990
Ridge with Ridge-GME	2704
Ridge with improved Ridge-GME (option 1)	2272
Ridge with improved Ridge-GME (option 2)	2205
Ridge with KS	2180
Ridge with HK	2171

The MSE for the Ridge Regression estimator using the improved Ridge-GME parameter estimator is slightly greater than the MSE for the Ridge Regression estimator based on the HK and KS estimates, although it is lower than the MSE for the OLS and the other Ridge Regression estimators presented in Table 3, including the Ridge Regression estimator using the original Ridge-GME parameter estimator. The Ridge Regression estimator using the improved Ridge-GME parameter estimator also provides a MSE less than the MSE of other Ridge Regression estimators considered by Muniz and Kibria (2009, pp. 628–629) and of some estimators considered by Sakallıoğlu and Kaçiranlar (2008, pp. 683–687) in the Portland cement model.

6 Conclusions

The Ridge-GME parameter estimator, by avoiding criticisms and difficulties of the traditional Ridge Regression literature and the GME estimation, seems to be a perfect combination that highlights the best of both methodologies. The Ridge-GME parameter estimator tries to solve the problem of the subjective selection of the tuning parameter by visual inspection of ridge trace in the traditional Ridge Regression procedure (by using the GME estimator) and the subjective selection of the supports for the parameters in the GME estimator (by using the ridge trace).

However, the Ridge-GME parameter estimator also requires the use of subjective information from the analysis of the ridge trace, which represents the main drawback of the estimator. The improvement accomplished in this paper overcomes this weakness of the original Ridge-GME parameter estimator. The simulation study and the empirical applications discussed in this paper, as well as the results obtained by Macedo et al. (2010) and Erdugan and Akdeniz (2012), reveal a good performance of this Ridge Regression procedure in the case of regression models with small samples sizes affected by collinearity.

Acknowledgements

I would like to express my gratitude to the referees. They offered extremely valuable suggestions for improvements. This work was supported by Portuguese funds through the CIDMA – Center for Research and Development in Mathematics and Applications, and the Portuguese Foundation for Science and Technology (“FCT - Fundação para a Ciência e a Tecnologia”), within project UID/MAT/04106/2013.

References

- Alkhamisi, M. A. and Shukur, G. (2007). A Monte Carlo study of recent ridge parameters. *Communications in Statistics - Simulation and Computation*, 36(3):535–547.

- Belsley, D. A., Kuh, E., and Welsch, R. E. (2004). *Regression Diagnostics - Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons, Hoboken, New Jersey.
- Brown, P. J. (1994). *Measurement, Regression, and Calibration*. Clarendon Press, Oxford.
- Caputo, M. R. and Paris, Q. (2008). Comparative statics of the generalized maximum entropy estimator of the general linear model. *European Journal of Operational Research*, 185(1):195–203.
- Ciavolino, E. and Al-Nasser, A. D. (2009). Comparing generalised maximum entropy and partial least squares methods for structural equation models. *Journal of Nonparametric Statistics*, 21(8):1017–1036.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407–499.
- Erdugan, F. and Akdeniz, F. (2012). Computational method for jackknifed generalized ridge tuning parameter based on generalized maximum entropy. *Communications in Statistics - Simulation and Computation*, 41(8):1411–1429.
- Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, 76(373):131–139.
- Golan, A. (2006). Information and entropy econometrics - a review and synthesis. *Foundations and Trends[®] in Econometrics*, 2(1-2):1–145.
- Golan, A., Judge, G., and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley & Sons, Chichester.
- Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223.
- Greene, W. H. (2008). *Econometric Analysis*. Pearson Prentice Hall, Upper Saddle River, New Jersey, 6th edition.

- Hald, A. (1952). *Statistical Theory with Engineering Applications*. John Wiley & Sons, New York.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition.
- Hoerl, A. E. and Kennard, R. W. (1970a). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1):69–82.
- Hoerl, A. E. and Kennard, R. W. (1970b). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975). Ridge regression: some simulations. *Communications in Statistics - Simulation and Computation*, 4(2):105–123.
- Jaynes, E. T. (2003). *Probability Theory - The Logic of Science*. Cambridge University Press, Cambridge.
- Kaçıranlar, S., Sakallıoğlu, S., Akdeniz, F., Styan, G. P. H., and Werner, H. J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on Portland cement. *The Indian Journal of Statistics, Series B*, 61(3):443–459.
- Khalaf, G. and Shukur, G. (2005). Choosing ridge parameter for regression problems. *Communications in Statistics - Theory and Methods*, 34(5):1177–1182.
- Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics - Simulation and Computation*, 32(2):419–435.
- Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics - Theory and Methods*, 32(5):1009–1020.
- Macedo, P., Scotto, M., and Silva, E. (2010). On the choice of the ridge parameter: a maximum entropy approach. *Communications in Statistics - Simulation and Computation*, 39(8):1628–1638.

- Maronna, R. A. (2011). Robust ridge regression for high-dimensional data. *Technometrics*, 53(1):44–53.
- Marquardt, D. W. and Snee, R. D. (1975). Ridge regression in practice. *The American Statistician*, 29(1):3–20.
- McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100.
- McDonald, G. C. and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350):407–416.
- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2006). *Introduction to Linear Regression Analysis*. John Wiley & Sons, Hoboken, New Jersey, 4th edition.
- Muniz, G. and Kibria, B. M. G. (2009). On some ridge regression estimators: an empirical comparisons. *Communications in Statistics - Simulation and Computation*, 38(3):621–630.
- Ryan, T. P. (2009). *Modern Regression Methods*. John Wiley & Sons, Hoboken, New Jersey, 2nd edition.
- Sakallıoğlu, S. and Kaçiranlar, S. (2008). A new biased estimator based on ridge estimation. *Statistical Papers*, 49(4):669–689.
- Satici, E. and Demirhan, H. (2012). Use of generalized maximum entropy estimation for freight flows modelling and an application. *Journal of Data Science*, 10:75–86.
- Stone, M. and Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society, Series B*, 52(2):237–269.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288.

- Woods, H., Steinour, H. H., and Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. *Industrial and Engineering Chemistry*, 24(11):1207–1214.
- Zhang, R. and McDonald, G. C. (2005). Characterization of ridge trace behavior. *Communications in Statistics - Theory and Methods*, 34(7):1487–1501.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67(2):301–320.